Translation of Manufacturer's Product Data for the ARROW Product Search System

Leonard Newnham and Robert Amor

Building Research Establishment

Bucknalls Lane, Garston, Watford WD2 7JR, UK

Phone +44-1923-664549, Fax +44-1923-664689

E-mail {newnhaml, trebor }@bre.co.uk

1 Abstract

ARROW (Newnham et al. 1997 and 1998) is a UK initiative that can provide access to any construction manufacturer's product information through virtual warehouses. This enables designers and specifiers to correctly identify products that match the parameters of their particular design. These warehouses rely on a system of indexing the manufacturer's product databases at regular intervals to keep the warehouse data up to date. To do this efficiently the manufacturers need to make their data available to the ARROW system in a standard format, which will probably be very different to the format of the manufacturer's internal database. For a manufacturer to participate in such a system there needs to be an easy and efficient method of translating product data information between different database structures.

The data model used by ARROW is an extended IFC model (Amor et al. 1998), and as such is a very complex structure when implemented with a relational database, (involving approximately 200 tables). To be accepted by manufacturers, the translation mechanism must take into account the following requirements. a) There is minimal re-entry of data. b) Once set up, it must be capable of running as an automatic process on the manufacturer's machine, so that as the manufacturer changes product data in the internal database so the data available to ARROW automatically changes without further user input. c) The full complexity of the IFC data structure is hidden from the user.

The approach taken by ARROW is to provide a toolkit for manufacturers that consists of; a) a flattened data structure, hiding the full complexity of the IFC data structure, and b) a tool for defining a mapping between the manufacturer's database fields and the flattened IFC structure.

This paper examines the particular problems such systems face in this area, and examines the approach used in the ARROW system to find a workable solution.

2 The ARROW Project

ARROW is a UK initiative that can provide access to any construction manufacturer's product information through virtual warehouses. This enables designers and specifiers to correctly identify products that match the parameters of their particular design. Benefits of this approach will be reduced abortive design work and reworking, a more reliable design specification, designs that can be constructed quicker, greater use of off-the-shelf components, savings in cost, and improved response time for customers. To make ARROW feasible a central search engine able to handle structured data and also free-text information has been developed. This is based on data collected from all manufacturers and suppliers willing to publish electronic information. This allows fast and accurate retrieval of specified product information, as well as delivering information in a form useable by CAD systems and other design tools (e.g., thermal simulation programs).

The main parts of the system comprise the following. First, the user, who has an interface with the system directly through a web browser or indirectly through a design tool. Second, the main system or Building Object Warehouse (BOW) server. Third, the distributed product databases at numerous sites across the Internet containing product data in a variety of formats. Finally, there is provision for interaction with existing KBS systems for product advice, design considerations, etc. Again, these can be distributed across the Internet. The main points are described briefly below and the system illustrated in fig 1.1.

User Environment The user can interact with the BOW system in one of two ways: either, directly through a web browser or alternatively, through a design tool. In the former case, the user interacts directly with the query handler. In the latter case the user may point to a wall and specify "window" with the design tool supplying the size parameters directly. An executable residing on the user's machine accepts parameters from the design tool and then interacts with the Query Handler on the user's behalf.

Query Handler The function of this is to help the user formulate exactly what he or she wants in terms of parameters used in the data model. The query handler contains knowledge of the product data model and so can prompt for parameters appropriate for a particular product type. There is provision for unspecified parameters, parameters within a range and exact values only, return sort order, etc. In addition, there is provision for free-text searches of the manufacturer's general product information, which is directly tied in to one query process.

Indexes, Indexers and the Search Engine The indexer has an authoritative list of all product repository addresses. Periodically these are searched, all product parameters retrieved and indexed. There are two indexes, one a structured database (currently a relational database) containing the minimum product information specified in the product model, the second for free text containing additional manufacturer information within html pages on a web server. The free text index is kept up to date using the Harvest indexer (Hardy et al, 1996).

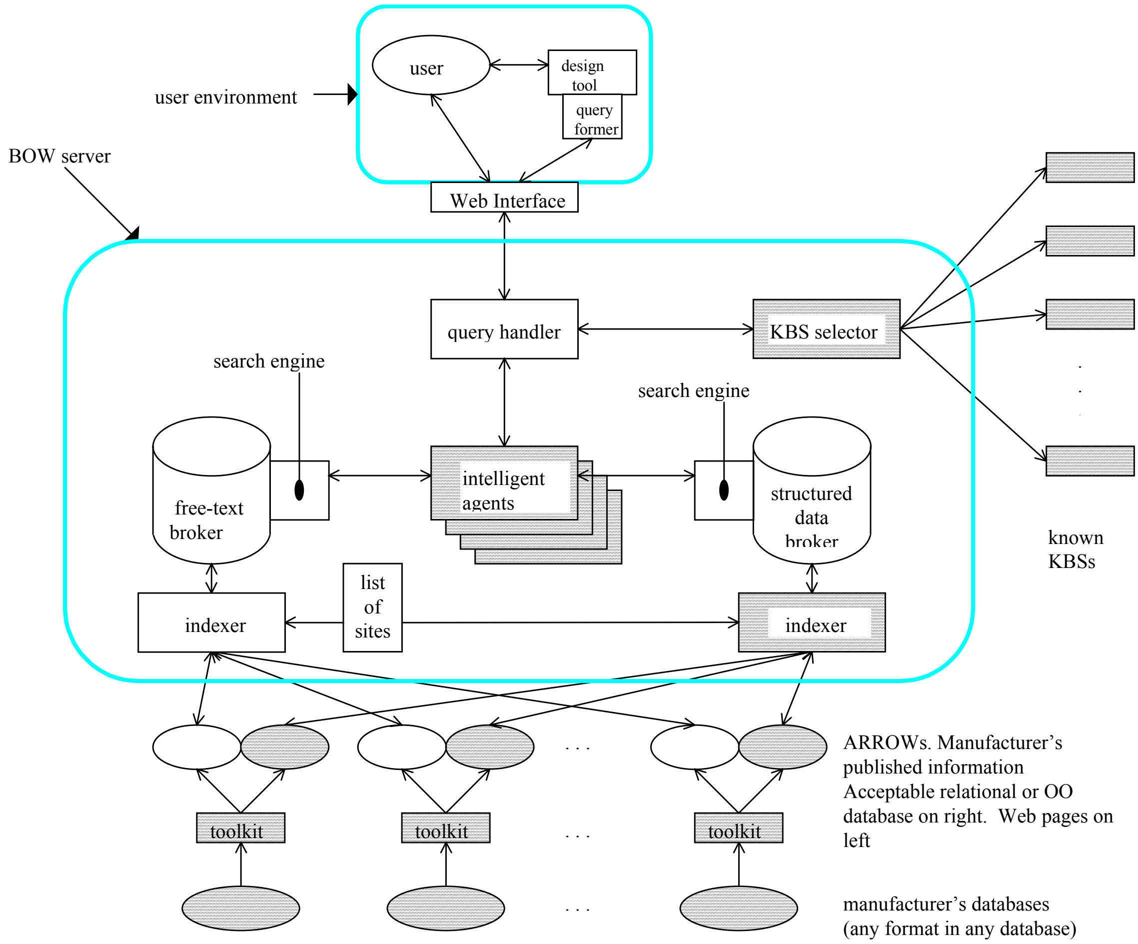

fig 1.1: BOW Architecture (shaded objects not implemented in demonstrator)

3 Data Model Representation Requirements

Analysing the requirements of the ARROW system, as described in section 1, led to a specification of the areas which needed model development. Analysis of the existing models showed a very low level of coverage of these areas in most developments, the exception being the IAI IFCs, which, while not being very detailed, gave the building blocks for almost all areas required. The final ARROW data model is therefore built entirely upon the IFCs, but with extensions to its structure to represent more specific areas. The main areas required to be covered were:

- Product and sub-assembly structures allow for the definition of all technical data about products. The IFCs give very basic information about building elements (e.g., door, window, distribution elements), which were specialised to contain all the parameters required for product selection. This part of the specification also defines the forms of graphical representation of products and the possible connections between related products, e.g., to ducts or walls, and through which mechanism. The product section also provides the links to all associated documentation about products, in terms of BBA certificates, manufacturer's specification sheets, CAD files of product detailing, VRML file, etc.

- Organisation structures allow for the specification of manufacturers and suppliers and the connections that exist between them for specific products. There is a triangular structure which joins a product to a manufacturer, and the manufacturer to a set of suppliers, and the supplier to a set of products (and vice-versa). The organisation structures provide the means for contacting manufacturers and suppliers as well as all information about their range of products and services.
- Catalogue information ties closely to products, manufacturers, and suppliers providing a mechanism to duplicate the type of catalogue services currently offered in the industry.
- Core information structures provide the majority of the basic information required in all other parts of the ARROW system. This includes: mechanisms to allow for multiple classifications for a single product; definition of materials and the layering of materials; specification of the cost for a product, and cost models which allow special relationships between suppliers and customers to be recognised; specification of standard units, and conversion factors for non-standard units; specification of measured values (e.g., luminosity, mass, volume); and documentation types to define whether a file is a CAD detail, or VRML, or a Word document and also whether it is a contract, detailed drawing, etc. This was the area which was best covered by the IFCs and required the fewest additions (the main exception being the document model specification).

The full specification of the ARROW data model can be found in Amor et al. (1998). This document describes all structures in the data model as well as routes to implementation of the data model. Mappings to common implementation forms are also provided, e.g., SQL, and Java.

{ARROW/IFC inheritance diagram here}

2 <u>Problems</u>

The aim of the toolkit is to assist the mapping from the manufacturer's database fields to the ARROW IFC data model. The translation process is a complex and difficult process, the toolkit will attempt to automate this to some degree.

There are several problems involved in the database translation process:

1. The semantics of the database fieldnames. The exact meanings that a database designer ascribes to a field within a database will vary according to the product represented. For example, three dimensions can be represented by length, breadth and height but clearly which is which depends on where the object is viewed from. A semantic knowledge of both databases is needed before meaningful translation can be carried out.
2. 1-to-1 mappings. These are straightforward once a pair of fields has been matched.
3. 1-to-many mappings. For example, full address → street, city, county and country. This is very difficult to automate as it requires knowledge of address structures to be able to identify each element successfully from the many different standard ways of writing an address.
4. Many-to-1 mappings. For example the reverse of (3). This is slightly easier but still needs to have an order imposed on the copying to ensure a sensible address is constructed.
5. 1-to-0 mappings, i.e. no obvious place to copy attribute to. This is bound to occur frequently as a manufacturer will wish to describe their product in a way that

differentiates it from the competition. Also new products will tend to have new features not catered for in older data models.

6. Complexity of the IFC data model. The path to reach much data for a product can be long, involving several intermediate tables. This makes any non-automated translation prone to error. Given there will be many cases like (3) above, data translation could become a complex, error prone and tedious task.

3 Research and Industry Developments

The UK government recently completed a scoping study for an industry knowledge base (IKB) for the construction industry (DETR 1996). This knowledge base is envisioned as a single point of entry to all information required by the construction industry, from news, journals, standards, codes of practice, practitioners and through to actual product information. The UK DETR (Department of Environment, Transport and Regions) have previously commissioned a demonstrator of such a system (Parand 1996) and are now supporting the concept of an IKB for the UK construction industry. The DETR has funded projects to examine various technical and commercial aspects of an IKB.

The concept of object oriented databases can exploit the work done on standards for exchange of product data (ISO STEP) (ISO 1994). The EC COMBINE project (Augenbroe 1995a and 1995b) has demonstrated how the STEP methodologies can be used to develop a single project database for sharing data between different design disciplines. A Building Components Database (BCD) (Parand 1995) was developed, by BRE, within the COMBINE project and linked with two architectural CAD systems across a network. This linkage allowed the project database to be populated using objects within the BCD. Hence demonstrating the potential for widespread data storage independent from particular applications. The ARROW project extends that work to encompass a selection of products from key suppliers and will encourage applications developers to use this database instead of each vendor developing their own, probably very functionally limited, database.

In the industry there is a shift in the medium used to publish manufactured product information. First with a switch from paper-based catalogues to CD-ROM based services. In the UK these services include: Glenigan Construction Database; Barbour Index; Technical Indexes; ASC Disk Index; Product Selector; and Product Selector Plus. Following on from this there are now several systems that are made available on the Internet, these include:

- The Building and Home Improvement Products Network (http://www.build.com/);
- First Source On-Line (http://www.afsonl.com/);
- Sweet's Group (http://www.sweets.com/);
- Certified Products (http://solstice.crest.org/index.shtml);
- SHARE Catalog Services Page (http://cdr.stanford.edu/html/SHARE/catalogs.html);
- ASC Web Index (http://www.ascwebindex.com/);
- Building Focus (http://www.building-focus.co.uk/)

A disappointment with these services is that in the main they duplicate the process required to use a paper catalogue, in that they use a simple classification scheme and don't enhance their systems through the inclusion of product parameters and associated files (e.g., CAD based standard details).

Recently there have been many national projects initiated in this area (for a comprehensive list see http://www.vtt.fi/cic/links/prodlib.html). These national projects demonstrate a higher level of functionality for users, in a similar manner to the UK ARROW project.

4 Approach Taken in ARROW

The ARROW project has tried to address the issues in section 4 as follows:

1. Provide a flatter database structure, considerably more simple than the full ARROW IFC structure. This will remove much of the unused parts of the full IFC structure and hence make the required junction tables and paths to data simpler. It will present the user with a more manageable data structure for the vast majority of products. Alongside this, ARROW will supply the full ARROW IFC data structure for those who wish to exploit the full data model.
2. For 1-to-1 and many-to-1 mappings ARROW is making an initial attempt towards table-based mapping. The user selects pairs of fields from the original and the flatter ARROW database, then copying occurs automatically.
3. Where there is no clear mapping the IFC structure has provision for addition of new fields. This is of use for specific attributes but as these fields will not be common to all products of that type then has the disadvantage of remaining invisible for the search process. Alternately, ARROW allows textual information to be stored on a web server for each product. It is easily searchable and thus preferable.
4. Once the manufacturer has entered all product information into the flatter ARROW database this is automatically translated to the full ARROW IFC database.

{db flow of information diagram here}

5 Conclusions

The development of an advanced manufactured product selection system requires the solution of many technical problems. It also requires solutions to many real-world interface issues. The work of ARROW presented in this paper discusses one of these problems, namely the interface between manufacturer information and an industry wide central information exchange. It is clear that manufacturer's can not be expected to support the large and complex data structures which are found in the comprehensive models developed for our industry. It is also clear that manufacturers hold a large amount of data required for all stages in a project, and an effective method of making that information available is of benefit to the industry. The ARROW approach to this problem is the provision of a toolkit for manufacturers which provides the following components. First, a flattened and simplified data structure for the manufacturer to enter information into. This data structure is close to the formats commonly found in current product catalogues. Second, a tool to allow manufacturers to perform simple mappings from their internal database to the simplified representation. Third, a method for manufacturers to enter additional differentiating information for their products. The provided toolkit then automatically converts all of this data into the more complex form used in the central server. This approach makes the ideal of an ARROW-like system feasible for a manufacturer from a time and cost point of view.

6 References

Amor, R., Newnham, L., Parand, F. and Nisbet, N. (1998) The ARROW Data Model Specification, BRE publication, Watford, UK, March 98.

Augenbroe, G. (1995a) COMBINE-2 Final Report, CEC-DGXII, Brussels

Augenbroe, G. (1995b) An Overview of the COMBINE Project, The First European Conference on Product and Process Modelling in the Building Industry, Dresden, Germany, Oct 94, Scherer, R. (ed.), Balkema, Rotterdam, pp. 547-554.

DETR (1996) Construct-IT: bridging the Gap, Scoping study for the Construction Industry Knowledge Base, Construction Sponsorship Directorate, Department of the Environment, HMSO, UK.

Hardy, D.R., Schwartz, M.F. and Wessels, D. (1996) Harvest User's Manual, Technical Report CU-CS-743-94, Department of Computer Science, University of Colorado, USA.

ISO (1994) ISO/DIS 10303-1:1994. Industrial Automation Systems and Integration – Product Data Representation and Exchange Part 1: Overview and Fundamental Principles.

Newnham, L., Parand, F., Amor, R. and Nisbet, N. (1997) The ARROW Framework for Object Warehouses, CIB W78 Workshop on Information Technology Support for Construction Process Re-Engineering, IT-CPR-97, Cairns, Australia, 9-11 July, pp. 319-328.

Newnham, L., Amor, R. and Parand, F. (1998) Gaining Quality Manufactured Product Information through ARROW, PDT Days, BRE, UK, 25-26 March, pp. 39-46.

Parand, F. (1995) Combine BCD Final report, CEC-DGXII, Brussels, Obtainable from BRE.

Parand, F. (1996) The Construction Information Gateway Demonstrator, CIBSE Journal, vol.